Survival models and health sequences

Peter McCullagh[*]

January 15, 2013

Abstract

Medical investigations focusing on patient survival often generate not only a failure time for each patient but also a sequence of measurements on patient health at annual or semi-annual check-ups while the patient remains alive. Such a sequence of random length accompanied by a survival time is called a survival process. Ordinarily robust health is associated with longer survival, so the two parts of a survival process cannot be assumed independent. This paper is concerned with a general technique—time reversal—for constructing statistical models for survival processes. A revival model is a regression model in the sense that it incorporates covariate and treatment effects into both the distribution of survival times and the joint distribution of health outcomes. It also allows individual health outcomes to be used clinically for predicting the subsequent survival time.

Keywords: failure time; interference; preferential sampling; quality-of-life; randomization; revival assumption; semi-revival time; time reversal; treatment effect

1 Survival studies

A survival study in one in which patients are recruited according to a well-defined protocol, and their health status monitored on a regular or intermittent schedule until the terminal event, here assumed to be fatal. Covariates such as sex and age are recorded at the time of recruitment, and, if there is more than one treatment level, the assignment is presumed to be randomized. In a simple survival study, the health status $Y(t)$ at time t is a bare-bones binary variable, dead or alive, and the entire process is then summarized by the length of time $T > 0$ spent in state 1, i.e. the survival time. In a survival study with health monitoring, $Y(t)$ is a more detailed description of the state of health or quality of life of the individual, containing whatever information—pulse rate, cholesterol level, cognitive score or CD4 cell count—is deemed relevant to the study. The goal may be to study the effect of treatment on survival time, or to study its effect on quality of life, or to predict the subsequent survival time of patients given their current health history.

[*] Department of Statistics, University of Chicago, 5734 University Ave, Chicago, Il 60637, U.S.A. E-mail: pmcc@galton.uchicago.edu

Survival studies with intermittent health monitoring are moderately common, and likely to become more so as health records become available electronically for research purposes. Within the past few years, several issues of the journal *Lifetime Data Analysis* have been devoted to problems connected with studies of exactly this type. For a good introduction, with examples and a discussion of scientific objectives, see Diggle, Sousa and Chetwynd (2006) or Farewell and Henderson (2010). Section 8 of van Houwelingen and Putter (2012) is recommended reading.

In practice, the patient's health status is measured at recruitment ($t = 0$), and regularly or intermittently thereafter while the patient remains alive. To emphasize the distinction between the observation times and observation values, each time is called an appointment date, the set of dates is called the appointment schedule; an appointment cancelled is a non-appointment, and cancellation is assumed to be uninformative. Apart from covariate and treatment values, a complete uncensored observation on one patient $(T, \mathbf{t}, Y[\mathbf{t}])$ consists of a survival time $T > 0$, an appointment schedule $\mathbf{t} \subset [0, T)$, and the health status measurements $Y[\mathbf{t}]$ at these times. To accommodate patients whose record is incomplete, a censoring indicator variable is also included. In that case, the censoring time is usually, but not necessarily, equal to the date of the most recent appointment.

A statistical model for a survival study is a family of probability distributions for the record of each patient, all three components included. At a minimum, therefore, it is necessary to model the survival time and the state of health jointly, and to consider how the joint distribution might be affected by treatment. Ordinarily, robust health is associated with longevity, but if both are affected by treatment, there is no guarantee that the two effects are in similar directions.

In the sense that the health status is measured over time on each patient, a survival study is a particular sort of longitudinal study. Certainly, temporal and other correlations are expected and must be accommodated. But the distinguishing feature, that each sequence is terminated by failure or censoring, gives survival-process models a very distinct character.

The goal of this paper is not so much to recommend a particular statistical model, as to suggest a general mathematical framework for the construction of survival-process models, permitting easy computation of the likelihood function and parameter estimates, and straightforward derivation of predictive distributions for individual survival times. For example, the paper has nothing to say on the choice between proportional hazards and accelerated lifetimes for accommodating treatment effects. Although we have reservations concerning time-evolving covariates, all standard survival models are acceptable within the framework. Nor has the paper anything to contribute to the choice between Bayesian and non-Bayesian methods of analysis; prior distributions are not discussed, so either approach can be used. Administrative complications of the sort that are inevitable in medical and epidemiological research will be ignored for the most part, so no attempt is made to provide a complete turnkey package. For example, the paper has little to say about how best to handle incomplete records other than to recognize that censoring and delayed entry are issues that must be addressed—again using standard well-developed methods. Since most of the computations needed for model fitting and parameter estimation are relatively standard and need not involve specialized Markov chain or Monte Carlo algorithms, detailed discussion of computational techniques is omitted. The emphasis is on statistical principles, strategies for model formulation, sampling, and the distinction between time-dependent variables and time-evolving variables in the definition of treatment effects.

2 Latent-variable models

There are numerous instances in the medical and biostatistical literature where investigations generate both successive measurements, such as CD4 lymphocyte cell counts on each patient, together with survival time (DeGruttola and Tu, 1994; Guo and Carlin, 2004; Fieuws, Verbeke, Maes and Vanrenterghem, 2008). Geriatric studies seldom focus exclusively on survival time, but tend to emphasize variables related to quality of life, such as overall physical and mental health, mobility, independence, memory loss, mental acuity, and so on. In the statistical literature, survival studies with health monitoring are called longitudinal studies with time-to-event data (Henderson, Diggle and Dobson, 2000; Tsiatis and Davidian, 2004; Wu, Liu, Yi and Huang, 2012). Although there are variations in model formulation and implementation, all authors are agreed on the need for a joint distribution covering both survival time and the progression of health outcomes.

Motivated by several examples of a medical nature, Tsiatis and Davidian (2004) and Diggle, Sousa and Chetwynd (2008) provide a good explanation of the ins and outs of joint modelling. The inevitability of an association—positive or negative—between the measured health outcome and the survival time is one of the key model components. Beginning with Wulfsohn and Tsiatis (1997), the standard modelling strategy uses a bivariate zero-mean temporal process $\eta_i = (\eta_{0i}, \eta_{1i})$, independent and identically distributed for distinct patients, which affects the health outcome directly and additively and also the force of mortality h_i in a component-wise manner. The longitudinal health variable, also called the outcome variable, is modelled as a temporal stochastic process,

$$Y_i(t) = \mu(t) + \eta_{0i}(t)$$

in which μ is the mean function. Given η, the survival time for patient i is the time to the first event in the Poisson process whose intensity at time t is

$$h_i(t) = h_0(t) \times \exp(\eta_{1i}(t) + x_i'\beta),$$

where $h_0(t)$ is an arbitrary intensity common to all patients, and $x_i'\beta$ is the covariate effect. For $\eta_1 \equiv 0$, this is the proportional hazards model (Cox, 1972). The dependence between η_0 and η_1 has the desired effect of inducing a dependence between the survival time and health outcomes (Henderson, Diggle and Dobson, 2000, section 2.2; Rizopoulos, 2000; Sweeting and Thompson, 2011).

Since $\eta_i(t)$ is not observed, (thus not included in the history $\mathcal{H}_t$ generated by the observations up to time t), some authors prefer to replace $\eta_{1i}(t)$ with $\alpha Y_i(t)$, effectively treating the longitudinal variable as a time-evolving covariate in the proportional hazards model (Tsiatis and Davidian 2004; Sweeting and Thompson, 2011). This is not entirely satisfactory because η_1 is ordinarily envisaged as a relatively smooth function whereas η_0 invariably has a white noise component.

The conventional strategy for model construction seems natural enough for recurrent non-fatal events, but probability distributions constructed in this way are extraordinarily complicated when applied to single-event survival data. Tsiatis and Davidian (2004, section 3) give an explicit 'likelihood function', admitting that its derivation from the assumptions is unclear. They point out that the latent-variable model defines health-status trajectories extending beyond death, and this fact alone invites counterfactual speculation, which they do not dismiss but wisely decline to embrace. Moreover, some authors distinguish between the observed health trajectory and the 'true' health trajectory (Guo and Carlin, 2004), whereas others on the logical positivist side of the philosophical spectrum consider such a distinction neither appropriate nor inappropriate, but simply meaningless.

What is missing from the latent-variable formulation is the recognition of the fact that health outcomes are observed on patients only while they are alive. This is a sampling restriction that can be incorporated by flatlining, i.e. by re-defining the *observable* outcome process as $Y'(t) = Y(t) \times I(T < t)$, so that $Y'(t) = 0$ for $t \geq T$. Alternatively, but not exactly equivalently, the observable process is the restriction of Y to the patient's lifetime, which is the random domain $(-\infty, T)$. Both restriction and flatlining imply that T is a function of the observable process, though not of its finite restriction $Y'[\mathbf{t}]$, so the need for a joint distribution becomes less clear.

In the simplest case where the latent process is Gaussian, $\eta_0 \sim \mathrm{GP}(0, K_0)$ on $\Re$, the distributions for each finite subset $\mathbf{t} \subset \Re$ are Gaussian:

$$Y[\mathbf{t}] \sim N(\mu[\mathbf{t}],\ K_0[\mathbf{t}]).$$

This is not to be confused with the distribution of the observable process $(\mathbf{t}, Y'[\mathbf{t}])$ at either a fixed or randomly generated set of appointment dates. For fixed $\mathbf{t}$, the implied restriction to survivors $\{i\colon T_i > \max(\mathbf{t})\}$ is an instance of preferential sampling (McCullagh, 2008; Diggle, Menezes and Su, 2010), and the distribution among survivors is not Gaussian. For random $\mathbf{t}$, the list of health records $Y'[\mathbf{t}]$ is a point in the space $\cup_{n\geq 0}\Re^n$ of finite-length real-valued sequences, which is not a vector space. In either case, $Y'[\mathbf{t}]$ is not Gaussian.

In order to avoid some of these difficulties, both philosophical and computational, the suggestion put forward in this paper is to approach the problem from a different angle—literally in reverse. Time reversal is mentioned briefly in section 8.3 of van Houwelingen and Putter (2012), so the idea is not entirely new. But its full power does not appear to have been explored or exploited.

3 Time reversal

3.1 The survival process

A survival process Y is a stochastic process defined for real t, in which $Y_i(t)$ is the state of health or quality of life of patient i at time t, usually measured from recruitment. In a simple survival process, the state space $\mathcal{R} = \{0, 1\}$ is sufficient to encode only the most basic of vital signs, dead or alive; more generally, the state space is any set large enough to encode the observable state of health or quality of life of the patient at one instant in time. Flatlining is the distinguishing characteristic of a survival process, i.e. $\flat \in \mathcal{R}$ is an absorbing state such that $Y(t) = \flat$ implies $Y(t') = \flat$ for all $t' \geq t$. Whatever the state space may be, a survival process clearly cannot be stationary.

Survival time is the time to failure:

$$T_i = \sup_{t \geq 0} \{t : Y_i(t) \neq \flat\};$$

it is presumed that $Y(0) \neq \flat$ at recruitment, so $T_i \geq 0$. This definition is quite general, and does not exclude immortality, i.e. $T = \infty$ with positive probability. In all of the models considered here, however, failure time is finite with probability one.

In constructing a probability distribution for the record of one patient, it is essential to bear in mind that the three observed components $(T, \mathbf{t}, Y[\mathbf{t}])$ cannot be independent. First, $T > 0$ is a positive random variable, and the appointment schedule $\mathbf{t}$ is a finite subset of the random interval $[0, T)$, which implies that $\mathbf{t}$ is also a random variable. Moreover $T > \max(\mathbf{t})$ implies that the appointment schedule for any patient is informative about his or her survival time. Likewise, since the number of health-status measurements coincides with $\#\mathbf{t}$, it also should be strongly correlated with survival. Second, if better health status is associated with longer survival, we should expect patients who are initially frail to have shorter health-status records than patients who are initially healthy. In other words, even if the trajectories for distinct individuals may be identically distributed, the first component $Y(0)$ of a short health-status record should not be expected to have the same distribution as the first component of a longer record. On the contrary, any model such that record length is independent of record values must be regarded as highly dubious for survival studies.

Despite these complications, we aim to construct a statistical model that has the right sorts of symmetries and is not self-contradictory or otherwise inappropriate in any of the senses discussed above. Progress in this direction requires a few assumptions, and in this paper, the first assumption is that

$$\mathbf{t} \perp\!\!\!\perp Y \mid T. \qquad (1)$$

In other words, given the survival time, the appointment schedule is independent of the patient's state of health. This condition does not require appointments to be made regularly or kept sedulously, but it does demand that the probability of an appointment being missed while the patient lives should not depend on the state of health, except through T. A similar assumption is made explicitly or implicitly by most authors: see Henderson, Diggle and Dobson (2000, section 2.1). The assumption is mathematically natural, but it is not one to be taken for granted.

3.2 The revival assumption

On the assumption that the failure time is finite, the time-reversed survival process

$$Z_i(s) = Y_i(T_i - s)$$

is called the revival process. Thus, $Z_i(s)$ is the state of health of patient i at time s prior to failure, and $Z_i(T_i) = Y_i(0)$ is the value at recruitment. By construction, $Z(s) = \flat$ for $s < 0$, and $Z(s) \neq \flat$ for $s > 0$. Although Z is defined in reverse time, the temporal evolution via the survival process occurs in real time: by definition, $Z(\cdot)$ is not observable component-wise until the patient dies.

The transformation $Y \mapsto (T, Z)$ is clearly invertible; it may appear trivial, and in a sense it is trivial. Its one key property is that the revival process Z and the random variable T are variation independent. In the statistical models considered here, variation independence is exploited through the revival assumption, which states that the revival process and the survival time are statistically independent. More generally, $Z \perp\!\!\!\perp T \mid X$ if covariates are present.

To understand what the revival assumption implies, consider two patients i, j with identical covariate values $x_i = x_j$, whose survival times are $T_i = 5$ and $T_j = 20$ time units respectively. Exchangeability implies that their health status values at revival time s, $Z_i(s)$ and $Z_j(s)$, are identically distributed, and the revival assumption implies that both are independent of the failure times. The revival assumption is trivially satisfied by survival models with simple follow-up, where $Y(t)$ is binary,

The chief motivation for time reversal and the revival assumption has to do with the effective alignment of patient records for comparison and signal extraction (Fig. 2). Are the temporal patterns likely to be more similar in records aligned by age (time since birth or recruitment), or are they likely to be more similar in records aligned by reverse age (time remaining to failure)? Ultimately, the answer must depend on the context, but the context of survival studies suggests that the latter may be more effective than the former. Figures. 8.3 and 8.4 of van Houwelingen and Putter (2012), which are not substantially different from Fig. 2 of this paper, offer strong confirmation of this viewpoint in at least one survival study involving white blood cell counts for patients suffering from chronic myeloid leukemia. For an application unrelated to survival, see example B of Cox and Snell (1981). Although it is mathematically natural, and perhaps even biologically reasonable, no guarantee can be offered that the revival assumption (independence) will be satisfied in any specific circumstance. Fortunately, the revival assumption can be checked, and certain sorts of departure can be accommodated.

Model construction by time reversal may appear peculiar and unnatural in biological work, where the accepted wisdom is that an effect such as death cannot precede its supposed causes, such as ill health. However, the contrarian viewpoint—that proximity to death is the chief cause of ill health—seems no less compelling and no more helpful. The author's attitude is that metaphysical discussion along such lines is seldom productive and best avoided.

3.3 Exchangeability

In the presence of covariates such as sex, age at recruitment or treatment status, exchangeability is understood in the sense of McCullagh (2008, section 2) i.e. it applies to each subset of patients having the same covariate value. For any such set of patients, it implies that the survival times $T_{i_1}, \ldots, T_{i_n}$ are identically distributed, the record lengths $\#\mathbf{t}_{i_1}, \ldots, \#\mathbf{t}_{i_n}$ are identically distributed, the health-status variables $Y_{i_1}(t), \ldots, Y_{i_n}(t)$ are identically distributed, and likewise for the revival values at any fixed revival time s. In this paper, therefore, baseline health status is the first component of Y, not a covariate. This is essential for revival models: it is immaterial that $Y(0)$ is measured prior to randomization and treatment assignment. Exchangeability does not imply that $Y_i(0)$ is independent of the record length $\#\mathbf{t}_i$.

Assuming that the record is complete, the observation for one patient consists of a survival time T, a finite appointment schedule $\mathbf{t} \subset [0, T)$, and a sequence of length $\#\mathbf{t}$ taking values in the space of medical records, here denoted by $\mathcal{R}$. For simplicity of notation in what follows, it is assumed that the appointment date is included in $\mathcal{R}$. Then the sample space for the observation $(T, Y[\mathbf{t}])$ on one patient is

$$\mathcal{S} = (0, \infty) \times \bigcup_{k=0}^{\infty} \mathcal{R}^k$$

in which the second component is the space of finite-length $\mathcal{R}$-valued sequences.

For n patients $i_1, \ldots, i_n$, the observations are independent if the joint distribution on $\mathcal{S}^n$ factors in the usual way:

$$P_{i_1,\ldots,i_n}(A_1 \times \cdots \times A_n) = \prod_{j=1}^{n} P_{i_j}(A_j),$$

where each P_i is a probability distribution on $\mathcal{S}$, and $A_1, \ldots, A_n \subset \mathcal{S}$ are arbitrary events. In that circumstance, it is sufficient to describe the marginal distributions P_i on $\mathcal{S}$, which may depend on covariates x_i. The observations on patients are infinitely exchangeable if $P_{i_1,\ldots,i_n}$ is the marginal distribution of $P_{i_1,\ldots,i_n,i_{n+1}}$, and all joint distributions are unaffected by permutation of patients, who are always assumed to be distinct individuals.

The implications of exchangeability are the same whether the record for each patient is expressed in terms of the survival process Y or the revival process. It implies that the revival processes are identically distributed. Together with the revival assumption, that Z and T are independent, it implies that $Z_{i_1}(s), \ldots, Z_{i_n}(s)$ are identically distributed independently of the survival times $T_{i_1}, \ldots, T_{i_n}$.

3.4 Covariates

In the absence of specific information to the contrary, responses for distinct units are presumed to be distributed exchangeably. In the great majority of situations, specific information does exist in the form of covariates or classification variables or relationships. A covariate is a function $i \mapsto x_i$ on the units, in principle known for all units whether they occur in the sample or not. A covariate implies a specific form of inhomogeneity such that equality of covariates implies equality of response distributions: $x_i = x_j$ implies $Y_i \sim Y_j$. In practice, approximate equality of x-values also implies approximate equality of distributions. Likewise, a relationship is a function on pairs of units such that $R(i, i') = R(j, j')$ implies $(Y_i, Y_{i'}) \sim (Y_j, Y_{j'})$ for distinct pairs provided that the two pairs also have the same covariate values: $(x_i, x_{i'}) = (x_j, x_{j'})$. Geographic distance and genetic distance are two examples of symmetric relationships. The overarching principle is that differences in distribution, marginal or joint, must be associated with specific inhomogeneities in the experimental material or observational units.

The status of certain variables in specific survival studies may appear genuinely unclear. The conventional rationalization, in which certain variables used for prediction are notionally 'fixed' or non-random and treated as covariates, is not especially helpful for survival studies. Consider, for example, marital status as one variable in a geriatric study in which the goal is to study both quality of life and survival time. However it is defined, quality of life is a multi-dimensional response, a combination of mobility, independence, optimism, happiness, family support, and so forth. Marital status is a temporal variable known to be associated with survival and with quality of life; one goal may be to predict survival given marital status, or even to recommend a change of status in an effort to improve the quality of life. Should such a variable be regarded as a covariate or as one component of the response? Another example of a similar type is air quality and its relation to the frequency and severity of asthmatic attacks (Laird, 1996). For survival studies, and for longitudinal studies generally, the answer is very clear and very simple: *every time-evolving variable is necessarily part of the response process.*

By definition, a temporal variable x is a function defined for every $t \geq 0$. A temporal variable is a covariate if it is also a function on the units, meaning that the entire function is determined and recorded at baseline. Usually this means that x is constant in time, but there are exceptions such as patient age: see also section 3.5. Marital status and air quality, however, are not only temporal variables, but variables whose trajectories evolve over real time; neither is available as a covariate at baseline.

With marital status as a component of the survival process, the joint distribution may be used to predict the survival time beyond t of an individual whose marital history and other health-status measurements up to t are given. For that purpose, it is necessary to compute the conditional distribution of T, or more generally of Y, given the observed history $\mathcal{H}_t$ up to time t. For such calculations to make mathematical sense, marital status must be a random variable, a function of the process Y. Thus, the statement 'marital status is a random process' is not to be construed as a sociological statement about the fragility of marriage or the nature of human relations; it is merely a mathematical assertion to the effect that probabilistic prediction is not possible without the requisite mathematical structure of σ-fields $\mathcal{H}_t \subset \mathcal{H}_{t'}$ for $t \le t'$ and probability distributions.

3.5 Treatment

Treatment refers to a scheduled intervention or series of interventions in which, at certain fixed or random times, the prescription for patient i is switched from one arm to another. Thus, $a_i(t)$ is the treatment arm scheduled for patient i at time $t > 0$, including $t > T_i$. In general, but crucially for revival models, a null level is needed for $t \le 0$, including the baseline $t = 0$. The entire temporal trajectory $a_i(t)$ for $t > 0$ is determined by randomization and recorded at $t = 0$. It does not evolve over real time in response to the doctor's orders or the patient's perceived needs, so it is not a time-evolving variable. In the sense that it is recorded at baseline, $a_i(\cdot)$ is a covariate; in the sense that it is a temporal function, it is a time-dependent covariate. Unlike a covariate, treatment is defined only for sampled units.

In practice, the distribution of $a(\cdot)$ is such that a switch of treatment arm is seldom scheduled more than once, and then only immediately after recruitment. Nonetheless, we maintain the more general formulation to underline the fact that treatment is a scheduled intervention such that $a_i(t) \ne a_i(0)$, and thus not constant in time. Unlike a survival process, the treatment schedule does not evolve in real time; it is contained in $\mathcal{H}_{0+}$, so the survival time is not a function of treatment.

Let $\bar{a}_i(s) = a_i(T_i - s)$ be the treatment arm expressed in revival time, so that, in the standard setting, $\bar{a}_i(s)$ is null for $s \ge T_i$. The revival assumption, that $Z \perp\!\!\!\perp T \mid \bar{a}$, may be regarded as trivial because T is a function of $\bar{a}$. In the case of treatment, however, the more important assumption is lack of interference, i.e. the treatment assigned to one individual has no effect on the distribution for other individuals, and the treatment assigned at one point in time has no effect on the response distribution at other times. For the latter, the statement is as follows. For each finite subset $\mathbf{s} \subset \Re^+$, the conditional distribution of $Z[\mathbf{s}]$ given the treatment schedule and survival time depends only on the treatment arms $\bar{a}[\mathbf{s}]$ prevailing at the scheduled times, i.e.

$$Z[\mathbf{s}] \perp\!\!\!\perp \bar{a} \mid \bar{a}[\mathbf{s}].$$

This is a strong assumption denying carry-over effects from earlier treatments or later treatments. It implies in particular that $Z(s) \perp\!\!\!\perp T \mid \bar{a}(s)$, which is primarily a statement about the one-dimensional marginal distributions.

It is common practice in epidemiological work for certain time-evolving variables to be handled as covariates, as if the entire trajectory were recorded at baseline. This approach is perfectly reasonable for a variable such as air quality in an asthma study where lack of cross-temporal interference might be defensible. It has the advantage of leading to simple well-developed procedures for effect estimation using marginal moments (Zeger and Liang, 1986; Zeger, Liang and Albert, 1988; Laird, 1996). The same approach does not make sense for an evolving variable such as marital status in a survival study, because the entire trajectory—suitably coded for $t > T_i$—would often contain enough information to determine the survival time.

4 Survival prediction

4.1 Simple Gaussian revival process

Consider first the simplest model in which observations for distinct patients are independent and identically distributed. To simplify matters further, problems related to parameter estimation are set aside. In other words, survival time T is distributed according to F, and the revival processes Z is distributed independently with known distribution G. The problem is to predict the survival time of an individual for whom the survival process at times $\mathbf{t} = (t_1 < \cdots < t_k)$ is given, $Y[\mathbf{t}] = y$. It is understood that $\mathbf{t}$ is a set of appointments, which implies that $Y(t_k) \neq \flat$, and hence that $T > t_k$.

To avoid a potential ambiguity of notation, it is worth re-stating the nature of the information provided for prediction. If $\mathbf{t}$ were known to be the complete appointment record, and if appointments were scheduled at unit intervals, we could reasonably deduce from the absence of subsequent appointments that $\max(\mathbf{t}) < T < \max(\mathbf{t})+1$. However, the information given is that contained in $\mathcal{H}(t_k)$, namely that the patient has had k appointments in the interval $[0, t_k]$, producing a partial record $Y[\mathbf{t}]$. Thus, no upper bound for T can be inferred.

For positive real numbers $\mathbf{s} = (s_1 > \cdots > s_k)$, let $g(z; \mathbf{s})$ be the joint density at $z \in \mathcal{R}^k$ of the heath status values

$$Z[\mathbf{s}] = (Z(s_1), \ldots, Z(s_k)) = (Y(T - s_1), \ldots, Y(T - s_k)).$$

Under the revival model, the joint density of $(T, Y[\mathbf{t}])$ at (t, y) is

$$f(t) \times g(y; t - \mathbf{t}) \qquad (2)$$

where $f = F'$ is the survival density. Then the conditional survival density given $Y[\mathbf{t}] = y$ is proportional to (2), which is zero for $t < \max(\mathbf{t})$.

If we were to include the schedule distribution $p(\mathbf{t} \mid T = t)$ as an additional factor in (2), we might consider a toy model such as a baseline appointment followed by a Poisson process with rate ρ on $(0, t)$, so that

$$p(\mathbf{t} \mid t) = \frac{e^{-\rho t}\, (\rho t)^{\#\mathbf{t}-1}}{(\#\mathbf{t}-1)!} \times \frac{(\#\mathbf{t}-1)!}{(\rho t)^{\#\mathbf{t}-1}} = e^{-\rho t},$$

where ρ is perhaps ten times the death rate. Because of this additional factor, the conditional survival density given $(\mathbf{t}, Y[\mathbf{t}])$ is very different from that derived in the preceding paragraph. This is precisely as it ought to be, because the latter calculation includes an additional critical piece of information—that $\max(\mathbf{t})$ is the patient's *final* appointment.

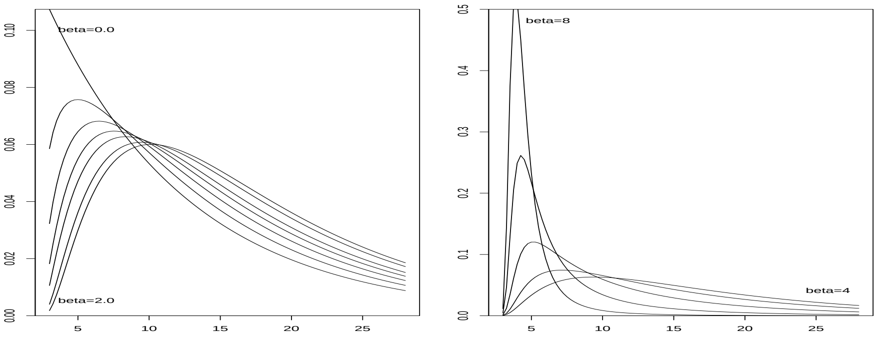

Figure 1: Conditional density of survival time for various values of β.

A simple numerical example illustrates the idea. Suppose T is exponentially distributed with mean 10 years, and the revival process for $s > 0$ is a real-valued Gaussian process with mean $E(Z(s)) = \beta s/(1+s)$ and covariance function $\delta_{ss'} + \exp(-|s - s'|)$ for $s, s' > 0$. The observed health-status values at $\mathbf{t} = (0, 1, 2, 3)$ are $y = (6.0, 4.5, 5.4, 4.0)$.

For $\beta = 0$, the conditional density is such that $T - 3$ is exponential with mean 10; for various values of β in the range $0 \le \beta \le 8$, the conditional density is shown in Fig. 1. Evidently, the conditional distribution depends on both the observed outcomes and on the model parameters: the median residual lifetime is not monotone in β. In applications where the mean function is estimated with appreciable uncertainty, the predictive distribution is an appropriately weighted convex combination of the densities illustrated.

4.2 The predictive density ratio

The ratio of the conditional survival density at t to the marginal density is proportional to the factor $g(y; t - \mathbf{t})$, in which $y, \mathbf{t}$ are fixed, and t the time. This modification factor—the Radon-Nikodym derivative—depends only on the revival process, not on the distribution of survival times. On a purely mathematical level, it is precisely the likelihood function in the statistical model for the k-dimensional variable $Y[\mathbf{t}]$ whose density at y is $g(y; t - \mathbf{t})$ for some value of the temporal offset parameter $t > t_k$.

In a Gaussian revival model, both the mean vector $\mu[t - \mathbf{t}]$ and the covariance matrix Σ of $Y[\mathbf{t}]$ may depend on t. In the simpler circumstance where Σ is independent of t, $\log g$ is a quadratic function of $y - \mu[t - \mathbf{t}]$, so the dependence of the likelihood on t stems from the lack of constancy of the mean function. Clearly g has its maximum at the value $\hat{t} \ge t_k$ that minimizes $\|y - \mu[t - \mathbf{t}]\|^2$ in the appropriate norm, which could occur at more than one internal point or at the extremes. For a locally stationary point, maximum or minimum, $\hat{\mu}'\Sigma^{-1}(y - \hat{\mu}) = 0$, where $\hat{\mu} = \mu[\hat{t} - \mathbf{t}]$ and μ' is the derivative.

In the very special case where $\mu(s) = \alpha + \beta s$ is linear in reverse time, and the covariances are independent of s for $s > 0$, the log density ratio factor

$$-\tfrac{1}{2}(y-\mu[t-\mathbf{t}])'\Sigma^{-1}(y-\mu[t-\mathbf{t}]),$$

is also quadratic in t. After substituting $\alpha+\beta(t-\mathbf{t})$ for the mean function, and expressing the log density ratio as a quadratic in t, it can be seen that the predictive density ratio at $t > \max(\mathbf{t})$ is the density at βt of the Gaussian distribution with mean

$$-\alpha+\mathbf{1}'\Sigma^{-1}(y+\beta\mathbf{t})\big/(\mathbf{1}'\Sigma^{-1}\mathbf{1}) = \bar{y}-\alpha+\beta\bar{\mathbf{t}}$$

and variance $1/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$. Ignoring the dependence on the data that comes from parameter estimation, the dependence of the predictive density ratio on the data for one patient comes through the weighted averages

$$\bar{y}=\mathbf{1}'\Sigma^{-1}y/(\mathbf{1}'\Sigma^{-1}\mathbf{1}),\qquad \bar{\mathbf{t}}=\mathbf{1}'\Sigma^{-1}\mathbf{t}/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$$

for this particular individual.

In applications to geriatric studies with Y representing some measure of physical health or mental acuity, it is reasonable to consider a revival model in which the mean $\mu(s)$ is monotone increasing and with an asymptote as $s\to\infty$. The inverse linear model $\mu(s)=\beta s/(\gamma+s)$ with asymptote β, and semi-asymptote $\mu(\gamma)=\mu(\infty)/2$, is a natural choice. In that case, the likelihood function is bounded as $t\to\infty$, so the tail behaviour of the predictive distribution is the same as that of the unconditional survival distribution.

4.3 Exchangeable Gaussian revival process

In a more general Gaussian model, the revival values for distinct patients are exchangeable but not necessarily independent. Revival models have much in common with plant growth-curve models in which $Z_i(s)=\mu+\eta_0(s)+\eta_i(s)$ is a sum of two independent zero-mean Gaussian processes, and the mean $\mu\equiv\mu(s)$ is constant across individuals, and possibly constant in time. Usually the common trajectory $\eta_0(\cdot)$ is moderately smooth but not stationary, perhaps fractional Brownian motion with $\eta_0(0)=0$. The individual deviations are independent and identically distributed and they incorporate measurement error, so $\eta_i(\cdot)$ is the sum of a continuous process and white noise. Thus, the Gaussian process is defined by

$$\begin{aligned} E(Z_{is}) &= \mu(s) \\ \operatorname{cov}(Z_{is},Z_{i's'}) &= K_0(s,s')+\delta_{ii'}K_1(s,s')+\sigma^2\delta_{ii'}\delta_{ss'} \end{aligned} \tag{3}$$

for some suitable covariance functions K_0, K_1, each of which can be expected to have a variance or volatility parameter and a range parameter. In the case of fractional Brownian motion, for example, $K(s,t)\propto s^\nu+t^\nu-|s-t|^\nu$ for some $0<\nu<2$, which governs the degree of smoothness of the random function.

For an new patient such that $Y[\mathbf{t}]=y$, the conditional survival density $\mathrm{pr}(T\in dt\mid \text{data})$ given the data, including the outcomes for the new patient, is computed in the same way as above. The second factor in (2) is the density at the observed outcomes of the Gaussian joint distribution whose means and covariances are specified above. This involves all $n+1$ patients including the new patient.

4.4 Illustration by simulation

Figure 2 shows simulated data for 200 patients whose survival times are independent exponential with mean five years. While the patient lives, annual appointments are kept with probability $5/(5+t)$, so appointment schedules in the simulation are not entirely regular. Health status is a real-valued Gaussian process with mean $E(Z(s)) = 10 + 10s/(10+s)$ in reverse time, and covariances

$$\mathrm{cov}(Z(s), Z(s')) = (1 + \exp(-|s-s'|/5) + \delta_{ss'})/2$$

for $s, s' > 0$, so there is an additive patient-specific effect in addition to temporal correlation. Values for distinct patients are independent and identically distributed. This distribution is such that health-status plots in reverse time aligned by failure show a stronger temporal trend than plots drawn in the conventional way. The state of health is determined more by time remaining before failure than time since recruitment. These trends could be accentuated by connecting successive dots for each individual, as in Fig. 2 of Sweeting and Thompson (2011), but this has not been done in Fig. 2.

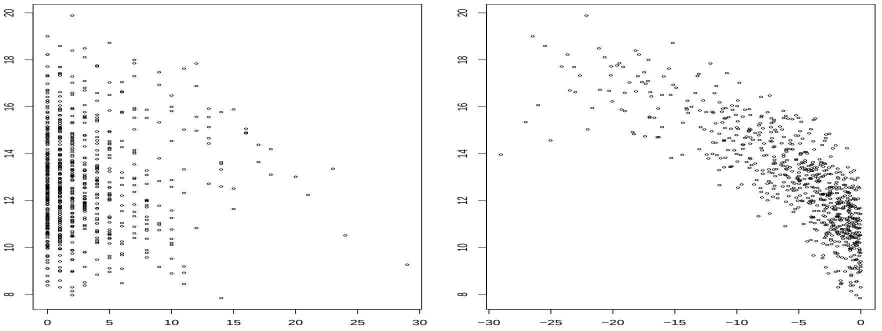

Figure 2: Simulated health status sequences aligned by recruitment time (left) and the same sequences aligned by failure time (right)

Since the survival times are exponential with mean five, independent of covariates and treatment, the root mean squared prediction error using covariates only is five years. For fixed $k \geq 2$, and a patient having at least k appointments, the conditional survival distribution given the first k health-status values has a standard deviation depending on the observed configuration, but the average standard deviation is about 2.5 years, and the root mean squared prediction error is about 2.7 years. For this setting, the longitudinal variable is a reasonably effective predictor of survival, and the prediction error is almost independent of k in the range 2–5. This summary does not tell the full story because certain y-configurations lead to very precise predictions whereas others lead to predictive distributions whose standard deviation exceeds five years.

The parameter settings used in this simulation may not be entirely representative of the range of behaviors of the conditional survival distribution given $Y[\mathbf{t}]$.

If the ratio of the between-patient to within-patient variance components is increased, the average variance of the conditional survival distribution decreases noticeably with k. For such settings, prediction using the entire health history is more effective than prediction using the most recent value.

4.5 Recurrent health-related events

In certain circumstances the outcome Y is best regarded as a point process, recording the occurrences of a specific type of non-fatal event, such as epileptic or asthmatic attacks or emergency-room visits. In other words, $Y_i \subset \Re$ is the set of times at which patient i experiences the event. Then $\mathbf{t} = (0, t_k)$ is a bounded interval, and the observation $Y[\mathbf{t}] = Y \cap \mathbf{t}$ is the set of events that occur between recruitment and the most recent appointment. This observation records the actual date of each event, which is more informative than the counting process $\#Y[(0,t_1)], \ldots, \#Y[(0,t_k)]$ evaluated at the appointment dates. If there are recurrent events of several types, Y is a marked point process, and $Y[\mathbf{t}]$ is the set of all events of all types that occur in the given temporal interval. The paper Schaubel and Zhang (2010) is one of several papers in the October 2010 issue of *Lifetime Data Analysis*, which is devoted to studies of this type.

We consider here only the simplest sort of recurrent-event process in which the revival process is Poisson, there is a single event type, and the subset $Y \cap \mathbf{t} = \mathbf{y}$ of observed event times is finite. The mean measure of the revival process is Λ, which is non-atomic with intensity λ on the positive real line. The density ratio at $t > \sup(\mathbf{t})$ is the probability density at the observed event configuration $t - \mathbf{y}$ as a subset of the reverse-time interval $t - \mathbf{t}$, i.e.,

$$g(\mathbf{y}; t - \mathbf{t}) = \exp(-\Lambda(t - \mathbf{t})) \prod_{y \in \mathbf{y}} \lambda(t - y).$$

In particular, if the intensity is constant for $s > 0$, the density ratio is constant, and the event times are uninformative for survival. In other words, it is the temporal variation of the intensity function that makes the observed configuration $\mathbf{y}$ informative for patient survival.

For a specific numerical example, let $\lambda(s) = (2+s^2)/(1+s^2)$ be the revival intensity, and let $\mathbf{t} = (0, 2)$ be the observation window. The revival intensity, monotone decreasing with an asymptote of one, implies that the recurrent events are moderately common at all ages, but their frequency increases as failure approaches. Figure 3 shows the likelihood as a function of $t \geq 2$ for three event configurations, $\mathbf{y}_0 = \emptyset$, $\mathbf{y}_1 = \{0.5, 1.2\}$ and $\mathbf{y}_2 = \{0.2, 1.3, 1.9\}$. Since the likelihood function is defined only up to an arbitrary multiplicative constant, the curves have been adjusted so that they are equal at $t = 20$, or effectively at $t = \infty$. In place of the predictive survival distributions, we show instead the ratio of the predictive hazard functions to the marginal hazards as dashed lines on the assumption that the marginal failure distribution is exponential with mean 5. Because of the form of the revival intensity, which is essentially constant except near the origin, the predictive hazard functions are very similar in shape to the likelihood functions.

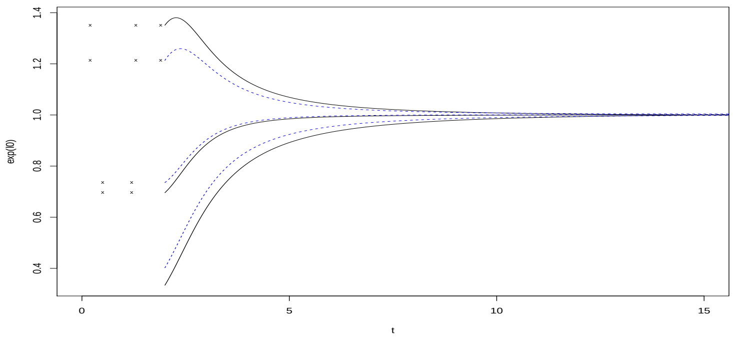

Figure 3: Likelihood functions for three point configurations (solid lines), with predictive hazard ratios (dashed lines)

5 Parameter estimation

5.1 Likelihood factorization

On account of independence, the joint density for the observations in a revival model factors into two parts, one involving only survival times, the other involving only the revival process. Although both factors may involve the same covariates and treatment indicators, the parameters in the two parts are assumed to be unrelated, i.e. variation independent. Thus the likelihood also factors, the first factor involving only survival parameters such as hazard modifiers associated with treatment and other covariates, the second factor involving only health-status parameters such as temporal trends and temporal correlations. In other words, the two factors can be considered separately and independently, either for maximum likelihood estimation or for Bayesian operations.

The first stage in parameter estimation is to estimate the survival distribution F together with treatment and covariate effects if needed. Whether the model for survival times is finite-dimensional or infinite-dimensional, this step is particularly simple because the first factor involves only the survival times and survival distribution. The standard assumption of independent survival times for distinct patients simplifies the problem even further. Exponential, gamma and Weibull models are all feasible, as is Cox's (1972) proportional hazards model. Censored records are handled in the standard way, for example by using the Kaplan-Meier estimator if there are no covariates. The literature on this topic is very large, and this paper has nothing further to add.

The second stage, which is to estimate the parameters in the revival process, is also straightforward, but only if all records are complete with no censoring. Serial dependence is inevitable in a temporal process, and there may also be independent persistent effects associated with each patient, either additive or multiplicative. Gaussian revival models are particularly attractive for continuous health measurements because such effects are easily accommodated with block factors for patients and temporal covariance functions such as those included in the simulation in Fig. 2.

Thus the second stage involves mainly the estimation of variance components and range parameters in an additive Gaussian model. One slight complication is that the revival process is not expected to be stationary, which is a relevant consideration in the selection of covariance functions likely to be useful. Another complication is that the health status may be vector-valued, $Y(t) \in \Re^q$, so there are also covariance component matrices to be estimated. If the covariance function is separable, i.e.

$$\operatorname{cov}(Z_{ir}(s), Z_{ir'}(s')) = \Sigma_{r,r'} K(s, s')$$

for some $q \times q$ matrix Σ, maximum-likelihood estimation is straightforward. But separability is a strong assumption implying that temporal correlations for all health variables have the same pattern, including the same decay rate, which may not be an adequate approximation. Nevertheless, this may be a reasonable starting point.

The second stage requires all health records to be aligned at their failure times. Accordingly, a record that is right censored cannot be properly aligned. The simplest option is to ignore censored records entirely in the second stage, on the grounds that their information content is limited, and the estimating equations based on complete records remain unbiased. The inclusion of censored records is thus more a matter of statistical efficiency than bias, and the information gained may be disappointing in view of the additional effort required.

5.2 Incomplete records

If we choose to include in the likelihood the record for a patient censored at $c > 0$, we need the joint probability of the event $T > c$, the density of the subset $\mathbf{t} \subset [0, c]$, and the outcome $Y[\mathbf{t}]$ at y. On the assumption that censoring is uninformative, i.e. that the distribution of the subsequent survival time for a patient censored at time c is the same as the conditional distribution given $T > c$ for an uncensored patient, the joint density is

$$\int_{t \geq c} f(t)\, p(\mathbf{t} \mid c, t)\, g(y; t - \mathbf{t})\, dt$$

on the space of finite-length records. The second factor, the density of the appointment dates as a subset of $[0, c]$ for a patient surviving to time $t > c$, is presumed not to depend on the subsequent survival time $t - c$, in which case it may be extracted from the integral. It may be reasonable to assume that the distribution of appointment schedules is known, for example if appointments are scheduled administratively at regular intervals, in which case the second factor may also be discarded from the likelihood. Since the survival probability $1 - F(c)$ is included in the first-stage likelihood, the additional factor needed in the analysis of the revival model is

$$\frac{1}{1 - F(c)} \int_{t > c} f(t)\, g(y; t - \mathbf{t})\, dt$$

in which $\mathbf{t}$ may regarded as a fixed subset of $[0, c]$. Unfortunately, the integral involves both the survival density $f(t) = F'(t)$ and the density of the revival process, so the full likelihood no longer factors. For an approximate solution, f may be replaced with the estimate obtained from the first-stage analysis of survival times.

For a revival parameter θ, a censored record contributes less information than an uncensored record. The log likelihood derivative generated by a complete record $(t, \mathbf{t}, y)$ is

$$U_\theta(t, y) = g'_\theta(y; t - \mathbf{t})/g_\theta(y; t - \mathbf{t})$$

where g'_θ is the derivative with respect to the parameter of the revival density. The log likelihood derivative for an incomplete record is the predictive expected value

$$\bar{U}_\theta(y) = \int_{t>c} U_\theta(t, y)\, f_\theta(t \mid y, \mathbf{t} \subset [0, c])\, dt.$$

Accordingly, the ratio of the observed Fisher information from an incomplete record to that from a complete record is

$$\frac{E(i_\theta(T, y)) - \mathrm{var}(U_\theta(T, y))}{i_\theta(t, y)},$$

where the mean and variance refer to the predictive survival distribution given $\mathbf{t} \subset [0, c]$ and $Y[\mathbf{t}] = y$. In general, unless the variance of the predictive distribution is small, this ratio will not be large.

5.3 Treatment effect: definition and estimation

Consider a survival study in which each eligible patient is randomized to one of two or more treatment arms $i \mapsto a_i$, which remains constant for $t > 0$. Health status is measured at recruitment—at $t = 0$, pre-randomization—and subsequently thereafter on a fixed appointment schedule until the patient dies. In addition to treatment and baseline health status, covariates x_i such as sex, and age (at recruitment) are also recorded: all covariates are constant in time.

We consider here only the simplest sort of revival model for the effect of treatment on patient health, ignoring entirely its effect on survival time. Health status in the revival process is assumed to be Gaussian, independent for distinct patients, and the treatment is assumed to have an effect only on the mean of the process, not on its variance or covariance. Consider two patients, one in each treatment arm,

$$a_i(t) = \bar{a}_i(T_i - t) = 1, \qquad a_j(t) = \bar{a}_j(T_j - t) = 0$$

such that $x_i = x_j = x$. The revival assumption asserts that the random variable $Z_i(s) - Z_j(s)$ is distributed independently of the pair T_i, T_j. By definition, the treatment effect as defined by the revival model is the difference of means

$$\tau_{10}(s; x) = E(Z_i(s)) - E(Z_j(s)) = E(Y_i(T_i - s)) - E(Y_j(T_j - s))$$

at revival time s. This is not directly comparable with either of the the conventional definitions

$$\gamma_{10}(t) = E(Y_i(t) - E(Y_j(t)) \quad \text{or} \quad \gamma'_{10}(t) = E(Y_i(t) - E(Y_j(t) \mid T_i, T_j > t)$$

in which the distributions are compared at a fixed time following recruitment. The expectation in a survival study—that healthy individuals tend to live longer than the frail—implies that $E(Y(t) \mid T)$ must depend on the time remaining to failure. In that case, the conventional treatment definition $\gamma'_{10}(t)$ depends explicitly on the difference between the two survival times. In other words, it does not disentangle the effect of treatment on patient health from its effect on survival time.

In the revival process, the treatment effect may be persistent in time, i.e. constant for $s > 0$ while the treatment is activated, but specifically excluding $s \geq \min(T_i, T_j)$, prior to randomization where the null level prevails. More complicated time-dependent effects are also possible, but will not be considered here. The treatment effect may also be equal in subsets of patients who have different covariate values, but interactions may also be present. For example, the treatment effect may be adverse for males but beneficial for females.

For a single patient surviving to time T, and an appointment schedule $\mathbf{t} \subset [0, T)$ equivalent to $\mathbf{s} = T - \mathbf{t}$ in reverse time, the revival outcome $Y[\mathbf{t}] = Z[\mathbf{s}]$ is distributed as

$$Z[\mathbf{s}] \sim N(\mu[\mathbf{s}],\ K[\mathbf{s}])$$

independently of T. For simplicity, the covariance function K is assumed to be independent of x and treatment. Assuming that the treatment effect is constant across individuals, i.e. independent of x, and also persistent over time, the revival mean has the form

$$\mu_i(s) = \alpha(s, x_i) + \tau(\bar{a}_i(s))$$

with a temporal trend α depending on x, and an additive treatment effect τ. Various other forms of dependence, some simpler and others more complicated, can be specified using factorial model formulae in the standard manner—after the series have been aligned in revival order.

Parameter estimates may be obtained by maximum likelihood using standard software for computing Gaussian likelihood functions. By contrast with the standard practice in the analysis of longitudinal studies, (...), it would be most unnatural in this setting to work with the conditional likelihood given the baseline outcomes $Y_i(0) \equiv Z_i(T_i)$. That is part of the reason for recommending that baseline response values not be treated as covariates.

6 Revival review

The easiest way to check the revival assumption is to formulate and fit a specific alternative model in which the revival process is not independent of the survival time. We consider here only the simplest design in which all records are complete, there are no covariates or treatment assignment, observations for distinct patients are independent, and the revival model is a family of Gaussian process. One way to do this is to replace (3) with

$$E(Z_i(s) \mid T) = \mu(s, T_i)$$

for some suitable family of functions $\mu(s, T)$, leaving the covariances unchanged. In particular, if x denotes patient age at recruitment, the revival mean might well depend on age at failure, $x + T$.

Consider, for instance, the non-linear Gaussian revival model with mean

$$\mu(s) = \alpha + \beta s/(\gamma + s),$$

which is such that $\mu(0) = \alpha$, $\mu(\infty) = \alpha + \beta$, and $\mu(\gamma) = \frac{1}{2}(\mu(0) + \mu(\infty))$, so that $\gamma > 0$ is the semi-revival time. Within this family, the revival trajectory for one patient could be different from that of another, depending on their survival times. In other words, α, β, γ could depend on T or $x + T$, either of which is a violation of the revival assumption. One of the simplest models of this type is the time-accelerated revival model in which the semi-revival time is inversely related to survival,

$$\mu(s, T) = \mu_0(sT) = \alpha + \beta sT/(\gamma + sT).$$

As a practical matter, it would be more effective to replace γ with $\gamma_0 + \gamma_1/T$ or $\exp(\gamma_0 + \gamma_1/T)$ to generate a test of the revival assumption. Likewise, we could replace α with $\alpha_0 + \alpha_1 T_i$, asserting that the outcome sequences for long-lived patients are elevated by a constant amount at all revival times. Similarly, if β is replaced with $\beta_0 + \beta_1 T_i$, the the asymptote is elevated in proportion to the additional lifetime.

Any modification of this sort is a violation of the revival assumption, so the survival time and the revival process are no longer independent. However, the factorization of the likelihood function remains intact, so the analysis remains relatively straightforward. For example, a likelihood ratio statistic to test the revival assumption can be constructed by fitting two nested models to the revival process, one satisfying the revival assumption, the other involving T.

Acknowledgements

I am grateful to S.M. Stigler for helpful comments.

7 References

Cox, D.R. (1972) Regression models and life tables (with discussion). J. Roy. Statist. Soc. B, 34, 187–220.

Cox, D.R. and Snell, E.J. (1981) Applied Statistics. London: Chapman and Hall.

DeGruttola, V. and Tu, X.M. (1994) Modeling progression of CD-4 lymphocyte count and its relation to survival time. Biometrika 80, 475-488.

Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994) Analysis of Longitudinal Data. Oxford Scienc Publications: Clarendon Press.

Diggle, P.J., Sousa, I. and Chetwynd, A. (2008) Joint modeling of repeated measurements and tome-to-event outcomes: The fourth Armitage lecture. Statistics in Medicine, 27, 2981–2998.

Diggle, P., Menezes, R. and Su, T-L. (2010) Geostatistical inference under preferential sampling (with discussion). Appl. Statist., 59, 191–232.

Fieuws, S., Verbeke, G., Maes, B. and Vanrenterghem (2008) Predicting renal graft failure using multivariate longitudinal profiles. Biostatistics 9, 419–431.

Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2004) Applied Longitudinal Data Analysis. New York: Wiley.

Guo, X. and Carlin, B. (2004) Separate and joint modeling of longitudinal and event time data using standard computer packages. American Statistician, 58, 1–10.

Henderson, R., Diggle, P. and Dobson, A. (2000) Joint modeling of longitudinal measurements and event time data. Biostatistics 1, 465–480.

Laird, N. (1996) Longitudinal panel data: an overview of current methodology. In *Time Series Models in Econometrics, Finance and Other Fields,* D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds. Chapman & Hall Monographs on Statistics and Applied Probability 65.

Lawless, J.F. and Nadeau, C. (1995) Some simple robust methods for the analysis of recurrent events. Technometrics 37, 158–168.

McCullagh, P. (2008). Sampling bias and logistic models (with discussion). J. Roy. Statist. Soc. B, 70, 643–677.

Molenberghs, G. and Verbeke, G. (2005) Models for Discrete Longitudinal Data. Springer.

Pepe, M. and Cai, J. (1993) Some graphical displays and marginal regression analysis for recurrent failure times and time-dependent covariates. J. Amer. Statist. Assoc. 88, 811–820.

Rizopoulos, D. (2010) JM: An R package for the joint modeling of longitudinal and time-to-event data. Journal of Statistical Software, 35, 1–33.

Sweeting, M.J. and Thompson, S.G. (2011) Joint modeling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical Journal 53, 750–763.

Tsiatis, A.A., DeGruttola, V. and Wulfsohn, M.S. (1995) Modeling the relationship of survival to longitudinal data measured with error., applications to survival and CD4 counts in patients with AIDS. J. Amer. Statist. Assoc. 90, 27–37.

Tsiatis, A.A, and Davidian, M. (2004) Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica, 14, 809–834.

van Houwelingen, H.C. and Putter, H. (2012) Dynamic Prediction in Clinical

Survival Analysis. Monographs on Statistics and Applied Probability 123; CRC Press.

Wulfsohn, M.S. and Tsiatis, A.A. (1997) A joint model for survival and longitudinal data measured with error. Biometrics 53, 330–339.

Zeger, S.L. and Liang, K.-Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121–130.

Zeger, S.L., Liang, K.-Y. and Albert, P. (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics 44, 1049–1060.